

Vertiv unterstützt Unternehmen bei der Überführung von KI-Pilotprojekten in den produktiven Einsatz

Ein neuer Leitfaden für die Implementierung, basierend auf Untersuchungen von Vertiv und STL Partners, hilft Unternehmen dabei, ihre KI-Archetypen zu bestimmen und einen geeigneten Standort für die erforderliche Infrastruktur zu ermitteln. Der Workloadbasierte Ansatz verknüpft Geschäftsanforderungen mit bedarfsgerechten Infrastrukturmodellen und Referenzdesigns, um Risiken bei der Bereitstellung zu minimieren.

München, 19. Oktober 2026 – Vertiv Holdings Co. (NYSE: VRT), ein weltweit führender Anbieter für kritische digitale Infrastrukturen, und das Deloitte Center for Integrated Research haben eine Zusammenarbeit vorgestellt, die Unternehmen dabei unterstützt, eine zentrale Herausforderung bei der Überführung künstlicher Intelligenz von der Testphase in den Produktivbetrieb zu bewältigen: die Entscheidung darüber, wo KI-Workloads ausgeführt werden sollen und ob die vorhandene physische Infrastruktur dafür geeignet ist.

Der neue „Vertiv Designing Infrastructure for Enterprise AI Guide“, entwickelt von Vertiv und STL Partners, einem Forschungsunternehmen für digitale Infrastrukturen, bietet Führungskräften aus den Bereichen Business, IT und Gebäudemanagement einen workloadbasierten Leitfaden für die Infrastrukturplanung. Ausgehend von den angestrebten Geschäftsergebnissen und dem jeweiligen KI-Workload können Unternehmen den Einsatzort, die Infrastrukturanforderungen sowie die Bereitschaft ihrer vorhandenen Infrastruktur bewerten. Anschließend können sie jedem Workload ein geeignetes physisches Infrastrukturmodell sowie ein Referenzdesign für die weitere Planung und technische Validierung zuordnen.

Die zugrunde liegende Untersuchung verdeutlicht einen entscheidenden Aspekt bei der Überführung von KI-Workloads in den dauerhaften Produktivbetrieb: Die Infrastruktur-Bereitschaft kann darüber entscheiden, wo Workloads bereitgestellt werden können und wie schnell sie produktiv genutzt werden können. Verfügbare Ressourcen in den Bereichen Stromversorgung, Kühlung, Konnektivität und Betriebskapazitäten bestimmen, ob ein Workload in einer bestehenden Einrichtung unterstützt werden kann oder ein anderes Bereitstellungsmodell erforderlich ist.

Der Leitfaden fasst diese Überlegungen in einem vierstufigen Prozess zusammen: (1) Definition des angestrebten Geschäftsergebnisses und des KI-Workloads, (2) Bewertung von Umfang, Leistungsanforderungen und Betriebsparametern des Workloads, (3) Auswahl des am besten geeigneten Einsatzorts sowie (4) Zuordnung von Workload und Bereitstellungsort zu einem physischen Infrastrukturmodell und einem Referenzdesign.

„Die nächste Entwicklungsphase der Unternehmens-KI wird nicht nur von Modellen und Rechenleistung geprägt sein, sondern auch davon, ob Unternehmen die erforderliche Stromversorgung, Kühlung, Konnektivität und Betriebsumgebung für die jeweiligen Workloads bereitstellen können“, sagte Giordano Albertazzi, CEO von

Vertiv. „Wenn Unternehmen beim jeweiligen Workload ansetzen, können sie Infrastrukturentscheidungen früher treffen, mögliche Engpässe rechtzeitig erkennen und einen belastbareren Weg von der KI-Vision zum produktiven Einsatz entwickeln.“

„Die wirtschaftlichen Rahmenbedingungen und betrieblichen Anforderungen von KI verändern sich, sobald Workloads in den produktiven Betrieb überführt werden“, sagte Nicholas Merizzi, Principal und U.S. Lead von Silicon2Service bei Deloitte Consulting LLP. „Unternehmen müssen Geschäftswert, Governance, operative Verantwortlichkeiten und Infrastruktur gemeinsam betrachten. Werden diese Aspekte frühzeitig in den Planungsprozess einbezogen, erhalten Entscheidungsträger eine fundiertere Grundlage für die Entscheidung, wo KI-Workloads ausgeführt werden sollten und wie ihr Betrieb langfristig sichergestellt werden kann.“

Der Ansatz trägt der Tatsache Rechnung, dass Unternehmens-KI aus einer Vielzahl unterschiedlicher Workloads besteht und nicht mit einer einzigen Infrastrukturentscheidung abgedeckt werden kann. Während explorative Workloads häufig in der Cloud verbleiben können, eignen sich für dauerhaft genutzte oder geschäftskritische Anwendungen eher hybride Infrastrukturen oder vom Unternehmen verwaltete Kapazitäten. Workloads, die kurze Reaktionszeiten, mehr Kontrolle, Datennähe oder eine bestimmte Modelleistung erfordern, müssen gegebenenfalls in Edge-Umgebungen oder in einer anderen kontrollierten Umgebung betrieben werden.

Der workloadspezifische Ansatz ermöglicht es, öffentliche Cloud-Angebote, Colocation-Rechenzentren, unternehmenseigene Kapazitäten, hybride Umgebungen und Edge-Infrastrukturen jeweils dort einzusetzen, wo sie den größten Mehrwert bieten. Anstatt sich pauschal für oder gegen die Cloud auszusprechen, bietet der Leitfaden eine konsistente Methode zur Bestimmung der geeigneten Betriebsumgebung für jeden einzelnen KI-Workload.

Die Studie ordnet KI-Workloads sechs Archetypen zu: Modellentwicklung, Retraining sowie vier Kategorien für den produktiven Einsatz. Diese Archetypen unterstützen Unternehmen dabei, Workloads mit ähnlichen Anforderungen an Leistung, Betrieb und Infrastruktur zu identifizieren und schaffen so eine einheitliche Grundlage für die Planung. Jeder Workload kann anschließend einem von vier physischen Infrastrukturmodellen zugeordnet werden: Edge AI Cluster, Small AI Cluster, Large AI Cluster oder High-Density AI Cluster.

Die vier Modelle bieten einen Ausgangspunkt für Entscheidungen in den Bereichen Stromversorgung, Kühlung, Steuerung, Konnektivität, Ausfallsicherheit, Betriebsmanagement und zukünftiges Wachstum. Die zugehörigen Referenzdesigns lassen sich auf Basis der Standortbewertung, der Workload-Anforderungen und der detaillierten technischen Planung anpassen.

Die Eignung der Infrastruktur ist bei allen vier Modellen von entscheidender Bedeutung. Der Leitfaden unterstützt Unternehmen dabei zu bewerten, ob ein vorgesehener Standort über ausreichende Kapazitäten in den Bereichen Stromversorgung, Kühlung, Konnektivität und Betrieb sowie genügend Spielraum für zukünftiges Wachstum verfügt. Eine frühzeitige Bewertung kann dazu beitragen, Infrastrukturengpässe zu identifizieren, bevor sie zu Hindernissen bei der Implementierung werden.

Der vollständige Leitfaden „Designing Infrastructure for Enterprise AI“ beschreibt die sechs Workload-Archetypen, die vier physischen Infrastrukturmodelle sowie eine vierstufige Roadmap für die Planung, Implementierung und Weiterentwicklung von KI-Infrastrukturen in Unternehmen. Er ist unter [Vertiv.com](https://www.vertiv.com)

verfügbar.

###

Über Vertiv

Vertiv (NYSE: VRT) bringt Hardware, Software, Analytics und Services zusammen, damit Applikationen bei Kunden kontinuierlich laufen, optimal performen und sich mit den Geschäftsanforderungen weiterentwickeln. Vertiv löst die wichtigsten Herausforderungen, denen Rechenzentren, Kommunikationsnetzwerke sowie kommerzielle und industrielle Anlagen ausgesetzt sind mit seinem Portfolio von Strom-, Kühlungs- und IT-Infrastrukturlösungen und Services, das Netzwerke von Cloud- bis Edge-Computing abdeckt. Vertiv ist in mehr als 130 Ländern weltweit tätig. Die Unternehmenszentrale befindet sich in Westerville, Ohio. Weitere Informationen und die aktuellsten Neuigkeiten und Inhalte von Vertiv finden Sie unter [Vertiv.com](https://www.vertiv.com).

Zukunftsgerichtete Aussagen

Diese Pressemitteilung enthält zukunftsgerichtete Aussagen im Sinne des Private Securities Litigation Reform Act von 1995, Abschnitt 27 des Securities Act und Abschnitt 21E des Securities Exchange Act. Diese Aussagen stellen lediglich eine Prognose dar. Tatsächliche Ereignisse oder Ergebnisse können erheblich von den Aussagen abweichen, die in dieser zukunftsgerichteten Aussage enthalten sind. Die Leser werden auf die bei der Securities and Exchange Commission eingereichten Unterlagen von Vertiv verwiesen, darunter der jüngste Jahresbericht auf Formblatt 10-K und alle nachfolgenden Quartalsberichte auf Formblatt 10-Q. Darin werden diese und andere wichtige Risikofaktoren im Zusammenhang mit Vertiv und dessen Geschäftstätigkeit erörtert. Vertiv ist nicht verpflichtet, seine zukunftsgerichteten Aussagen zu aktualisieren oder zu ändern, sei es aufgrund neuer Informationen, zukünftiger Ereignisse oder aus anderen Gründen, und lehnt diese Verpflichtung ausdrücklich ab.

Kontakt

Sascha Hommer
PR Director IT

riba:businesstalk

T +49 (0)261-963 757-24

E shommer@riba.eu